Decision rules and correlated features

Wahida Chowdhury (wahida_chowdhury@carleton.ca)
Institute of Cognitive Science, Carleton University
Ottawa, ON, K1S 5B6, Canada

Warren Thorngate (warren_thorngate@carleton.ca)
Department of Psychology, Carleton University
Ottawa, ON, K1S 5B6, Canada

Abstract

Most research on decision making attends more to cognitive rules than to the situations in which these rules are employed. One characteristic of these situations is the correlation among the features or cues they reveal. We simulated thousands of decision situations that varied in 1) the number of alternatives, 2) the number of features, and 3) the correlations among features. Six, simple decision rules were then used to choose an alternative and their choices were compared to those generated by a mathematically optimal rule. Results show that all rules, including some that do very poorly when features are uncorrelated, greatly increase their chances of making optimal decisions as feature correlations rise. We discuss some implications of these results for the use of simple decision rules in the real world.

Keywords: decision making; cognitive processes; mental rules; situations; simulation.

Varieties of Decision Rules

Thanks to the nature of our basic cognitive processes, humans have remarkable capacities to invent, store, employ, modify and share thousands of heuristics -- *mental rules* analogous to software applications -- to adapt to our environments. Many of these rules assist decision making, including those expressed by phrases such as "Do what you did before," "Ask your doctor," and "Flip a coin," and those documented in research on decision rules and multi-attribute choice (e.g., see Tversky & Kahneman, 1974; Yoon & Hwang, 1995).

Why do people use some rules more often than others? What are the consequences of employing the rules they use? Research reveals that rule choice is influenced by complex relationships among time, attention, memory and situational constraints, and by habit, motivation and emotion (for example, see Goldstein & Hogarth, 1997; Simon, 1956). Research also reveals that many of the most popular rules contain simplistic decision rules that ignore information potentially useful for good choices (Kahneman, 2011; Thorngate, 1980). As a result, bad decisions are made more frequently than necessary.

Still, simple rules do not always lead to bad decisions. Although some simple rules frequently lead to poor choices, others lead to mathematically optimal choices most of the time (Gigerenzer, Todd & ABC Research Group, 1999; Thorngate, 1980). For example, a simple rule prescribing that equal weights should be assigned to all features of several alternatives produces the same choice as a sophisticated, weighted-average model about 60-90% of the time (Thorngate, 1980).

Some differences in the capabilities of decision rules to generate optimal decisions stem from the amount and kind of information the rules employ and ignore. Other differences, however, are likely to stem from characteristics of the decision situation (see, for example, Chowdhury & Thorngate, 2013).

One situational characteristic is the relationship among features of alternatives. Consider, for example, an employer scanning application forms to select a new employee. If the forms contain questions about the age, education, work experience, and current salary of applicants, then the answers are likely to be correlated; younger applicants would probably have less education, work experience, and salary than would their older competitors. The correlations could benefit an employer who uses a decision rule that ignores most of the available information. If, for example, the intercorrelations among the answers averaged $r = +0.99$, then the employer's cursory evaluation of answers to only one question -- say, age -- would do almost as well as an attentive evaluation of answers to all questions. Under these conditions, even simplistic decision rules would likely do well.

How much intercorrelation is needed to compensate for simplistic decision rules? To answer the question we examined how (a) the correlations among features of decision alternatives, and (b) the nature of the decision rule employed to choose among these alternatives, influenced (c) the probability that the best alternative would be chosen. We also examined how these influences were moderated by (d) the number of alternatives available, and (e) the number of features describing each alternative.

We wrote a computer programme with functions that simulated seven decision rules believed to be common in making everyday choices. Each of these simulated rules was given thousands of decision situations that varied in the number of their alternatives and features, and in the average correlation among the features. One of the seven rules, the Max Weighted Rule, duplicated a classic, economic prescription for maximizing expected value. Different numerical weights were assigned to represent the importance of different features, and these weights were multiplied by the numerical values an alternative had on the corresponding features. A weighted average was then calculated to summarize the value of each alternative, and

the alternative with the highest weighted value was chosen and recorded. The computer programme then determined which alternative would be chosen by each of the remaining six decision rules, and tallied how often these choices matched the choices made by the Max Weighted Rule.

Method

The simulation was written in the *Julia* programming language (http://julialang.org), a recently-developed alternative to Matlab®. Each run of the simulation (1) created a decision situation where a simulated chooser must select one of several alternatives presented simultaneously, (2) applied six simplified decision rules to select an alternative, and (3) reported how often the simplified rules made the same choice as a sophisticated, max weighted-average rule.

Decision Situations

Each decision situation was represented by a matrix of specified number of columns and rows; each row in the decision matrix represented an alternative that could be chosen, and each column represented a feature the alternatives had. Each cell of the matrix was filled with a value for the corresponding alternative-feature combination. The feature-values were randomly generated from a normal distribution with an arbitrary mean = 100 and standard deviation = 15.

Each feature was also assigned a random weight (W) simulating its importance. The weights summed to 1.0 and were ordered so that the first column of a decision matrix had the most weight and last column of the matrix had the least.

Table 1 shows an example of a decision matrix with 3 alternatives, each with 3 features.

Table 1: Decision matrix.

Alternative	Feature 1 W=0.48	Feature 2 W=0.39	Feature 3 W=0.13
1	77.9	96.3	106.5
2	99.7	98.9	89.8
3	127.3	91.9	92.9

Independent Variables. We ran our simulation with many possible combinations of three independent variables. The first independent variable was the number of alternatives from which users could choose: 6, 12, or 24. The second independent variable was the number of features that each alternative had: 4, 8, or 16. The third independent variable was the average correlation between all pairs of features: r = +0.0, +0.2, +0.4, +0.6, +0.8, and +1.0.

Correlated values of the features were generated by a simple algorithm. We generated a normally-distributed random variable, Z, then generated values of features A, B, C, etc. by correlating each feature with Z. This resulted in samples of A, B, C, etc. that were correlated with each other with a standard mathematical formula: r(ac) = r(ab)*r(bc). If the population correlation between, say, A and Z was r = 0.5, between B and Z was 0.5, and between C and Z was 0.5, then our algorithm would generate samples of A, B and C that, on average, had correlations of r^2 = +0.25.

We ran the simulation for each of the 3*3*6 = 54 combinations of the three independent variables (three levels of number of alternatives * three levels of number of features in an alternative * six levels of correlations between features). For each combination, our programme simulated a specified set of alternatives with new, randomly generated numbers and weights for features. Then the programme chose alternatives using seven different decision rules and recorded these choices.

Decision Rules

Max Weighted (MW) Rule. The simulation programme multiplied feature values of each column by its corresponding weight and summed weighted values across each row to determine max-weighted alternative. For example, Feature 1 of each alternative in Table 1 was multiplied by its weight 0.48, Feature 2 was multiplied by 0.39, and Feature 3 by 0.13. These three weighted values were then summed to obtain a weighted average for each alternative. Then the MW rule chose the alternative with highest max-weighted average (Alt 3 = 109.05).

Equally Weighted (EW) Rule. The equally weighted decision rule ignored the given weights and instead assigned an equal weight to each feature. The rule then calculated a simple average feature-value for each alternative, and selected the alternative with highest simple average (see Coombs, Dawes, & Tversky, 1970, Chapter 5). In the example of Table 1, the EW rule would select third alternative because its equally-weighted average (.33*127.2 + .33*91.9 + .33*92.99 = 104) was highest of all three alternatives.

Highest Minimum (HMin) Rule. Similar to the classic minimax criterion of choice (see Coombs, Dawes, & Tversky, 1970, Chapter 5), the HMin decision rule first identified the minimum feature-value of each alternative. Then it selected the alternative with the highest of these minimum values. Weights assigned to features were ignored. In the example of Table 1, the HMin rule would select third alternative because its minimum feature-value (91.9) was highest of the three.

Highest Maximum (HMax) Rule. The HMax decision rule mirrored the HMin rule. It looked at maximum feature-value of each alternative and selected the alternative with the highest of these maximum values. Here too weights assigned to features were ignored. In the example of Table 1, the HMax rule would select the third alternative because its maximum feature-value (127.3) was higher than the other two maximums (106.5 and 99.7).

Max Most-Important (MMI) Rule. The max most important rule made rudimentary use of feature weights. It considered only the most heavily-weighted feature (column 1) of each alternative, selecting the alternative with highest value on this most-weighted feature. In the example of Table 1, the MMI rule would select third alternative because its feature-value (127.3) for the most weighted feature was higher than the values of other alternatives.

Max Least-Important (MLI) Rule. The max least-important decision rule was included as a whimsical variation of the MMI rule. It selected the alternative with highest feature-value for the *least* weighted feature. In the example of Table 1, the MLI rule would select the first alternative because its feature-value (106.5) for the least weighted feature was higher than the corresponding values (89.8 and 92.9) of the two other alternatives.

Lexicographic (L) Rule. This rule also made rudimentary use of weights assigned to features. Each alternative was first evaluated with respect to most-weighted feature (column 1). The average value of this feature (column 1 average) was computed, and alternatives with less-than-average values on the first feature were eliminated from further consideration. After elimination, if there were more than two alternatives remaining, those remaining were evaluated with respect to the second most-weighted feature (column 2). The average value of this feature (column 2 average) was then computed, and alternatives with less than the average value for the second feature were eliminated. This procedure was repeated until only one alternative remained. If there was a tie during elimination, one alternative was chosen at random. In the example of Table 1, the L decision rule would select the third alternative after eliminating alternatives with less than average value (101.7) for first feature (column 1).

Dependent Variables

The programme was iterated 1000 times for each of the 54 combinations of independent variables. When 1000 iterations of a combination were completed, the programme printed the percentage of trials on which the six decision rules (EW, HMin, HMax, MMI, MLI, and L rules) selected the alternative chosen by the max weighted (MW) rule.

Results

Figures 1-4 show the relations between correlations among features (x) and percentage of choices matching those of the MW decision rule (y), for the six other decision rules. Each figure displays one of the nine combinations of number of alternatives and number of features. The remaining five combinations produced similar results.

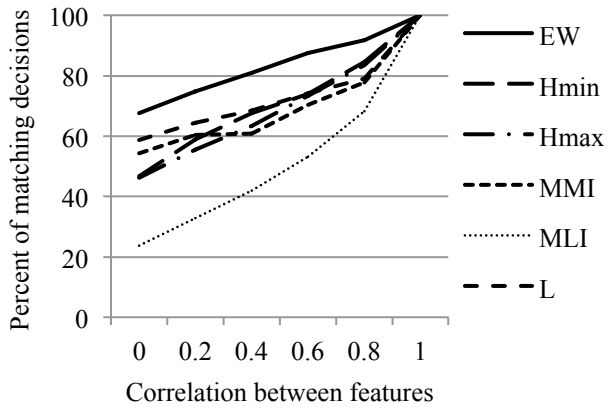

Figure 1: Percent of decisions matching those made by the MW rule in the 6-alternative and 4-feature condition.

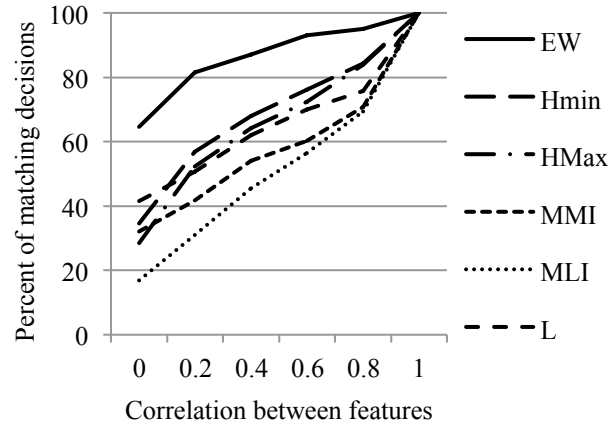

Figure 2: Percent of decisions matching those made by the MW rule in the 6-alternative, 16-feature condition.

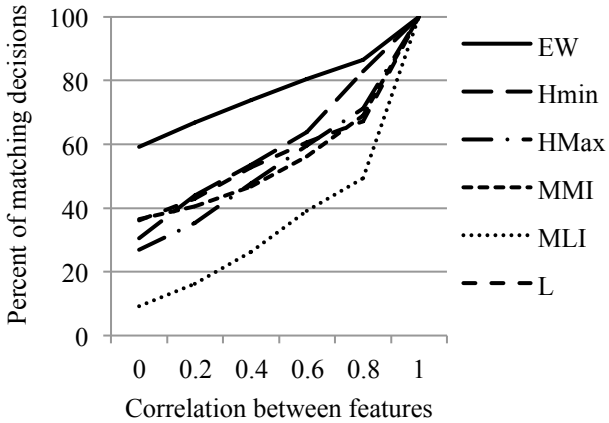

Figure 3: Percent of decisions matching those made by the MW rule in the 24-alternative, 4-feature condition.

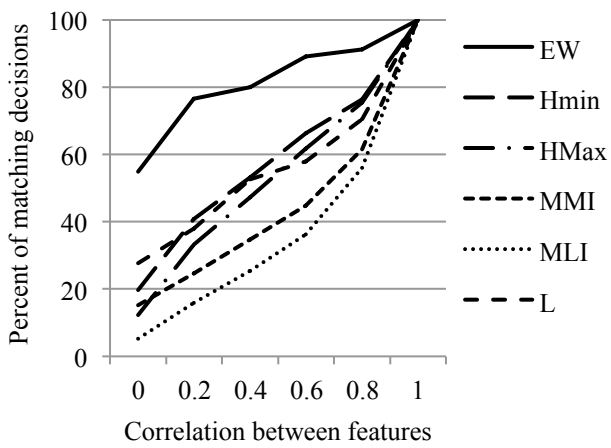

Figure 4: Percent of decisions matching those made by the MW rule in the 24-alternative, 16-feature condtion.

Figures 1-4 reveal three interesting results. First, when the features were uncorrelated, there was considerable variation in ability of different cognitive rules to produce the same results. The equal-weighted (EW) rule was most likely to select MW alternative across all nine research conditions, replicating the results of Thorngate (1980). Not surprisingly, the max-least-important rule (MLI) was the least likely to duplicate MW choices. The lexicographic (L) rule was the worst of the remainder, again replicating Thorngate (1980).

The second interesting result can be seen by comparing the family of curves across all four figures. Adding alternatives and features to the decision situation reduced the percent of choices replicating those of the MW rule. Adding alternatives produced a somewhat larger reduction in this percent than did adding features (compare, for example, the drop in percent shown in Figures 2 and 4 versus the drop in percent shown in Figures 1 and 3).

Perhaps the most interesting result, however, is the consistent rise in the percent of decisions matching those of the MW rule with the rise in feature correlations. Even moderate correlations of r = +0.4 to +0.6 produced increases of 20% or more in the number of times a rule mimicked the MW choice. As correlations approached r = +1.0, all rules mimicked the choices of the MW rule quite well and, of course, when all correlations reached r = +1.0, all rules generated the MW choice all the time.

Discussion

Popular, simple-minded decision-making rules are often criticized for increasing the chances of bad decisions (Gigerenzer, Todd & ABC Research Group, 1999; Thorngate, 1980). Our simulation shows that these chances decline when the features describing decision alternatives are correlated. When the features are uncorrelated, the choice of a decision rule can make a big difference in the chances of a suboptimal choice. When the features are highly correlated, almost any simple decision rule will do.

The conclusion prompts at least two questions. First, how correlated are the features of decision situations in the world? Second, should they be? Consider a judge assigned the task of choosing the winner of a competition for an academic scholarship based on information about the applicants' high school marks. Chances are the marks are correlated; marks in Grade 10 math, for example, are likely to be highly correlated with marks in Grade 11 math. Under these conditions, it might be reasonable for the judge to save time and mental effort by focusing on Grade 11 mark and ignoring Grade 10 marks, since it is likely the resulting choice would be the same as one made by considering the latter.

Alas, the correlations among features of alternatives in real decision situations are rarely known. The results of our simulation suggest that more attention be paid to them. Knowledge of the correlations would give decision makers a sense of the consequences of choosing a simple-minded decision rule.

Such knowledge, however, prompts a more vexing question: Should decision situations be constructed to present correlated features, or to ensure the features are correlation free? Most people would likely judge it unfair, for example, to construct a scholarship application form with the following two questions:

1. What was your average mark in Grade 10?
2. What was your average mark in Grade 10?

Yet the same people might judge the application form fair if "Grade 10" in Question 2 were replaced by "Grade 11," even though the Grade 11 information would likely be almost the same. The perfect application form might contain only items highly correlated with scholarly merit but uncorrelated with each other. Could such items be found? And would their utility be vitiated whenever a simple-minded decision rule is used? Answers to these questions are beyond the scope of the current simulation. But they deserve to be addressed in future studies.

References

Chowdhury, W., & Thorngate, W. (2013). Take it or leave it: Cognitive rules for satisfying choices. In R. West & T. Stewart (eds.), *Proceedings of the 12th International Conference on Cognitive Modeling*, Ottawa: Carleton University.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: an elementary introduction.* Englewood Cliffs, New Jersey: Prentice-Hall.

Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple rules that make us smart*, 563-607.

Goldstein, W., & Hogarth, R. (Eds.) (1997). *Research on judgment and decision making: Currents, connections and controversies*. Cambridge: Cambridge University Press.

Julia programming language. (2013). http://julialang.org.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus, and Giroux.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review; Psychological Review, 63(2),* 129.

Thorngate, W. (1980). Efficient decision rules. *Behavioral Science, 25(3),* 219-225.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Rules and biases. *Science, 185(4157),* 1124-1131.

Yoon, K. P. & Hwang, C-L. (1995). *Multi-attribute decision making: An introduction*. Thousand Oaks, CA: Sage Publications.